Enabling Semantic Interoperability

for Earth Science Data

Rob Raskin, Michael Pan, Chris Mattmann
Jet Propulsion Lab

Objectives

- Enable various Earth science-related projects to interoperate semantically
  - Provide a common "namespace" to define science and data-related terms
- Develop large collection of ontologies for Earth Sciences
  - Primarily land-ocean-atmosphere system

What is Semantic Interoperability?

- Two people/software/agents mean the same thing when referring to the same term
  - Lossless communication
- Namespace provides means to "register" a name
- Ontology provides means to clarify the definition of a term in a namespace

Semantics Enables Shared Understanding of Concepts

- Provides a namespace for scientific terms…plus
- Provides descriptions of how terms relate to one another
- Enables object in "data space" to be associated formally with object in "science concept space"
- "Shared understanding" enables software tools to find "meaning" in resources
- Described well using XML

Knowledge Reuse

- Semantic Web for Earth and Environmental Technology (SWEET) is an implementation of semantic interoperability
- SWEET is a concept space
- Enables scalable *classification* of Earth science concepts
  - Search engines use Open directory to classify contents of WWW space
- For educational use, SWEET supports navigable discovery of Earth science concepts, such as for an electronic encyclopedia
- Concept space is translatable into other languages/cultures using “sameAs” notions

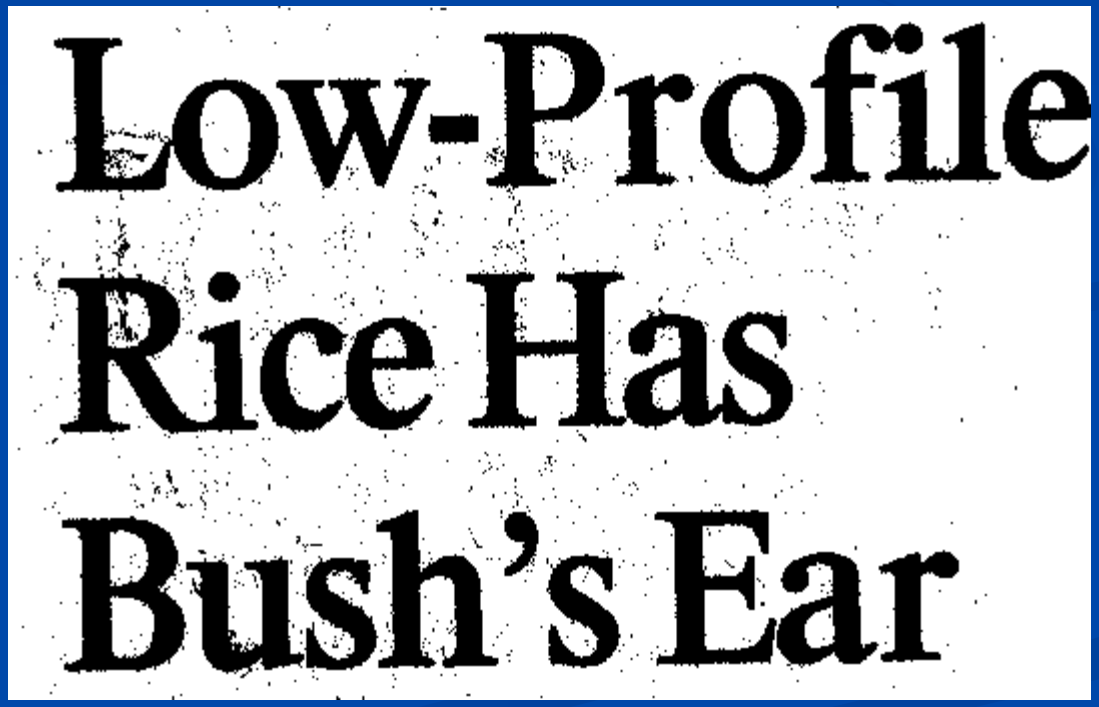
Semantics to Improve Search: Common Challenges

- Same word with multiple meanings
  - Cook (Captain)
  - Cook (chef)
  - Cook (county in Illinois)
- "Wrong" (synonymous) term
  - Marine vs. sea
- Too general or too specific
  - California vs. Los Angeles

Semantic Understanding is Difficult!

Let's eat, Grandma.
Let's eat Grandma.

Time flies like an arrow.
Fruit flies like a pie.



Ontology vs. Taxonomy

- Ontology
  - Children are subclasses of parent concepts
    - Parent properties inherited by children
      - multiple inheritance generally supported
  - Scalable
    - "New" concepts are often definable using multiple inheritance (e.g. Sea floor temperature) rather than creating a new definition
- Taxonomy
  - Not all properties are inherited

Barriers to Semantic Interoperability

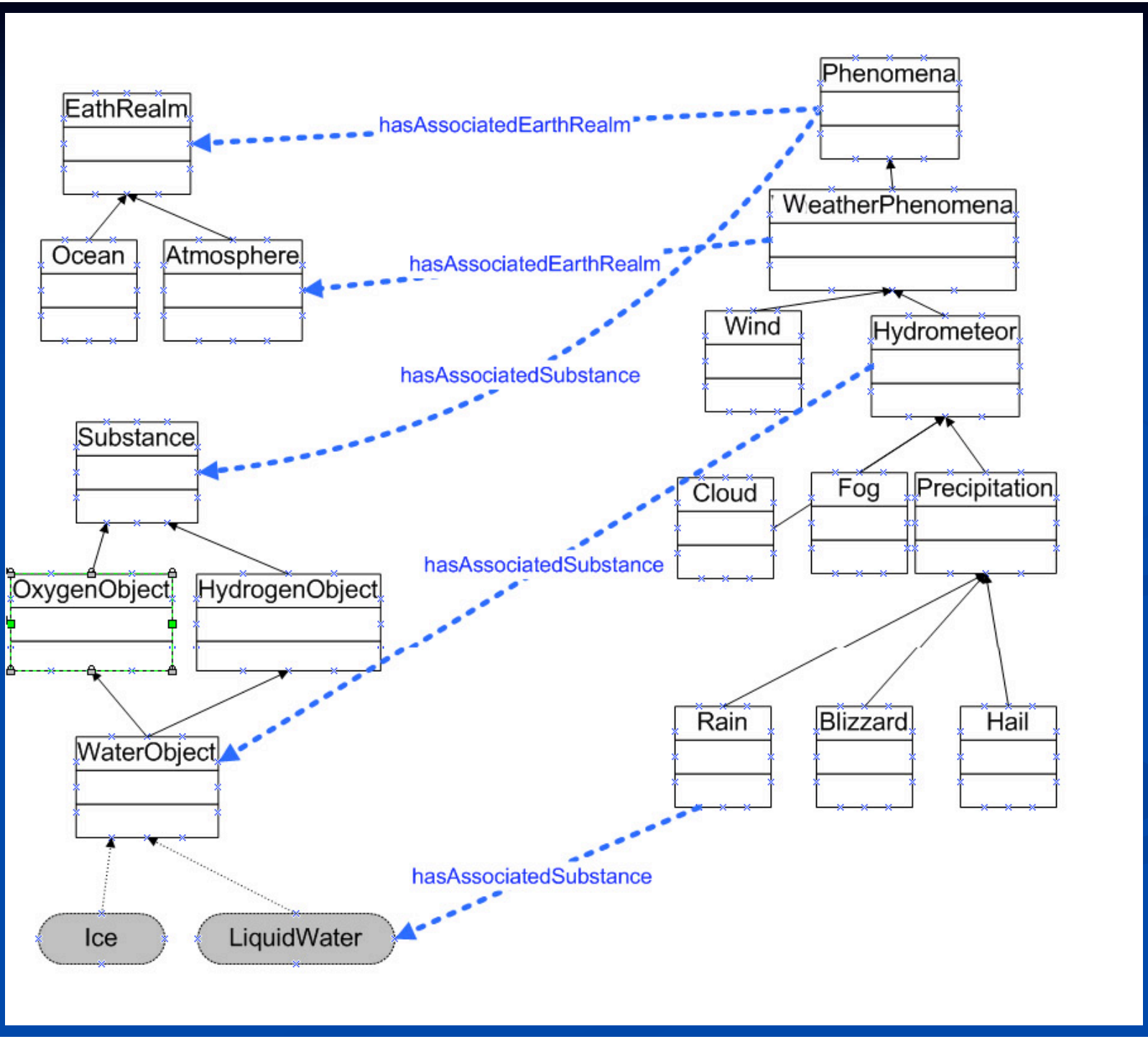
- We tend to not see beyond our local "communites of interest" for purposes of sharing knowledge
- Why take the effort to add semantic metadata when it is "obvious" what is meant?
- Why create a formal namespace when the meaning of the terms are "obvious"?

Semantic Interoperability Using Standard Languages

- Use standard ontology language to enable ontologies to import contents from one another
- Specialized domains can extend existing ontologies

Ontology Representation in OWL

- Ontology Web Language (OWL) accepted by the World Wide Web Consortium (W3C) as a language standard
- Several XML semantic languages defined, each a proper subset of next:
  - RDF (Resource Description Framework)
  - OWL Lite
  - OWL DL (Description Logic)
  - OWL Full
- OWL is outgrowth of DAML+OIL
  - DARPA no longer part of language development


Phenomena
EathRealm
hasAssociatedEarthRealm
WeatherPhenomena
Ocean
Atmosphere
hasAssociatedEarthRealm
Wind
Hydrometeor
hasAssociatedSubstance
Substance
Cloud
Fog
Precipitation
hasAssociatedSubstance
OxygenObject
HydrogenObject
Rain
Blizzard
Hail
WaterObject
hasAssociatedSubstance
Ice
LiquidWater

SWEET Ontologies

- SWEET includes two types of ontologies
  - Orthogonal (faceted) concepts:
    - Earth Realm (atmosphere, ocean, ...)
    - Property (temperature, pressure, CO2 conc., ...)
    - Substance (air, water, ash, ...)
    - Space (Antarctica, tropics, Washington, ...)
    - Time (Winter, Jan 2004, …)
  - Unifying concepts:
    - Phenomena (hurricane, earthquake, …)

SWEET Ontologies (cont.)

- Taken together, the two types of ontologies mirror the two types of scientific investigation:
  - Reductionism
  - Holism

Data Ontology

- Ancillary Data Interpretation
  - Units
    - Associate standardized units with a data object
  - Scale factor/offset
  - Missing value
- Models
  - Inputs, outputs, parameters
- Grid concepts
- Data services
  - Subset, regrid, …

ESML Integration

- ESML provides syntax tags
- SWEET ontologies complement ESML by providing a namespace for semantic tags
- Greatly expands the possible semantic tags in ESML

Translations

- CF Names (used by ESMF)
  - Long names such as: air_temperature_at_top_of_boundary_layer
  - Conversion table maps these concepts to SWEET ontology terms (adding new concepts, as needed)
- GCMD
  - Conversion table maps these concepts, decomposing complex concepts into faceted concepts

Incorporating External Knowledge

- Knowledge in typical on-line knowledge bases is too large and too dynamic to store locally
  - Gazetteers, earthquake databases
- We created OWL wrappers for selected database contents
  - Getty and CIA gazetteers
  - USGS Earthquake database

Incorporating External Knowledge (cont.)

- We created a general schema for creating wrappers for a large class of on-line resources
- We created a wrapper for WMS interface specification to access georeferenced mapped information, on demand
  - Wrapper creates URL with appropriate parameters

Spatial Ontology – Polygon Representation

- All major gazetteers use bounding box to to describe extent of a region
  - Size would be too large
- We used ancillary geographic databases with polygon boundaries to represent country and state boundaries as polygons
  - Bounding box of California does not include Nevada!

Spatial Ontology – Polygon Representation (cont.)

- Storage of polygon boundaries in PostGres DBMS
  - Native polygon data type
  - But does not handle polygon with large number of edges
  - Requires generalization (resolution reduction)

Mathematical Ontology

- Expanded mathematical ontology includes:
  - Multidimensional concepts
  - Cartesian products
- Using these concepts, we defined spatial concepts and spatial operators
  - These concepts can be used to carry out spatial reasoning in future projects

Search Engine

- Developed an expanded search engine that eliminates need for exact term match
  - Finds alternative search terms, then submits union of terms to GCMD Search Engine
- Alternative terms can be
  - Synonyms
  - Parent
  - Child

Challenges

- Difficulty in getting participants to see value of common semantics
- Semantic web still far from reality, due to limited generic tools for entering ontologies and exploiting contents
- Difficulty in visualizing large ontologies
  - It is very difficult to gain community review of ontologies

What Next?

- ESIP Federation support
  - Incorporation into Federation Interactive Network for Discovery (FIND)
    - Federation Search Tool
  - Ontologies maintained by the Federation
    - Products & Services Committee
  - Community review at August Federation Meeting in Asheville

Contacts

- SWEET http://sweet.jpl.nasa.gov

- Rob Raskin raskin@jpl.nasa.gov
- Mike Pan mjpan@jpl.nasa.gov